Cell PRESS
Structure
Article

Early Folding Events Protect Aggregation-Prone Regions of a β-Rich Protein

Ivan L. Budyak,[1,3] Beena Krishnan,[1,3,5] Anna M. Marcelino-Cruz,[2,4] Mylene C. Ferrolino,[1] Anastasia Zhuravleva,[1] and Lila M. Gierasch[1,2,*]
[1]Department of Biochemistry and Molecular Biology
[2]Department of Chemistry
University of Massachusetts Amherst, Amherst, MA 01003, USA
[3]These authors contributed equally to this work
[4]Present address: Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA
[5]Present address: G.N. Ramachandran Protein Center, Institute of Microbial Technology, Council of Scientific and Industrial Research, Chandigarh 160036, India
*Correspondence: gierasch@biochem.umass.edu
http://dx.doi.org/10.1016/j.str.2013.01.013

SUMMARY

Protein folding and aggregation inevitably compete with one another. This competition is even keener for proteins with frustrated landscapes, such as those rich in β structure. It is interesting that, despite their rugged energy landscapes and high β sheet content, intracellular lipid-binding proteins (iLBPs) appear to successfully avoid aggregation, as they are not implicated in aggregation diseases. In this study, we used a canonical iLBP, cellular retinoic acid-binding protein 1 (CRABP1), to understand better how folding is favored over aggregation. Analysis of folding kinetics of point mutants reveals that the folding pathway of CRABP1 involves early barrel closure. This folding mechanism protects sequences in CRABP1 that comprise cores of aggregates as identified by nuclear magnetic resonance. The amino acid conservation pattern in other iLBPs suggests that early barrel closure may be a general strategy for successful folding and minimization of aggregation. We suggest that folding mechanisms in general may incorporate steps that disfavor aggregation.

INTRODUCTION

The unavoidable competition between protein folding and aggregation links the energy landscapes of these two processes (Clark, 2004; Hartl et al., 2011; Jahn and Radford, 2008; Vendruscolo, 2012). However, in large measure, we lack adequately detailed descriptions of these landscapes to gain insights into vulnerabilities of proteins to aggregation and how to reduce the inherent risks. Predominantly β sheet proteins are a particularly apt structural class for examining molecular details of the folding/aggregation balance, as they have rugged folding energy landscapes with a high degree of frustration and consequent population of intermediate states and increased susceptibility to aggregation (Chavez et al., 2004; Plaxco et al., 1998). Nonetheless, there are several large families of β-rich proteins with no association with amyloid diseases, arguing that mechanisms exist to enable them to favor folding over aggregation. Some protection from aggregation may be attributable to chaperone action, but we hypothesize that intrinsic folding mechanisms must also incorporate strategies to minimize aggregation.

Intracellular lipid-binding proteins (iLBPs), a large family of successful folders (Banaszak et al., 1994), offer an excellent model system to explore this hypothesis. Despite their complex folding landscapes and extremely diverse sequence space, these 10-stranded β-barrel-containing proteins (Figure 1A) have not been implicated in any misfolding diseases. This observation suggests that iLBPs have evolved a robust folding mechanism with built-in aggregation protection. Here, we searched for topology-determining structural motifs as signatures of this mechanism. We used a model iLBP, cellular retinoic acid-binding protein 1 (CRABP1). Its rugged folding landscape is characterized by two on-pathway intermediates—an early collapsed state (I_1) and a later β-molten globule-like state with native topology (I_2)—and thus exemplifies β-barrel frustration (Clark et al., 1997; Clark et al., 1998).

The β barrels of iLBPs are formed via several medium-range contacts between strands and long-range contacts that link N- and C-terminal sequences to close the barrel (Figure 1A). The major conserved hydrophobic core of iLBPs comprises several highly networked conserved long-range interactions between the front and back sheet (Banaszak et al., 1994; Gunasekaran et al., 2004; Kleywegt et al., 1994; Marcelino et al., 2006) (orange spacefill in Figure 1B). In addition, there is a smaller cluster of primarily hydrophobic interactions that was identified in a study of conserved pairwise interactions (green spacefill in Figure 1B) (Gunasekaran et al., 2004). This minor core comprises portions of the helix-turn-helix motif that caps the β barrel, β strands 1′ and 10, and turns II and IV. It is intriguing that the minor core involves interactions between two local structural motifs, helix-turn-helix and turn IV, that have been shown (along with turn III) to populate native-like conformations in peptide studies (backbone in red, Figure 1B) (Rotondi and Gierasch, 2003b; Sukumar and Gierasch, 1997). This propensity to form

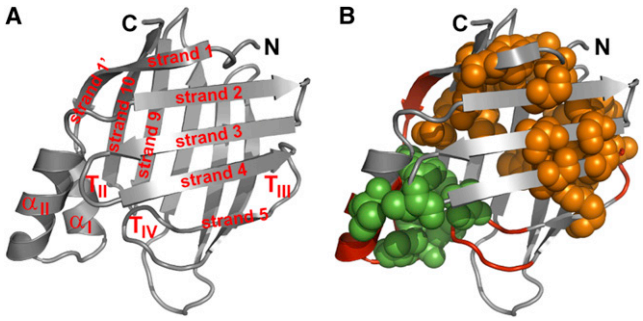

Figure 1. Structural and Topological Features of CRABP1

(A) Secondary structure elements of CRABP1 (shown on the structure of holo-CRABP1, Protein Data Bank ID 1CBR, with ligand omitted; the structure was visualized using PyMol software; W.L. DeLano, http://www.pymol.org). Helices are designated as α_I and α_{II}, and turns II, III, and IV are designated as T_{II}, T_{III}, and T_{IV}, respectively. Strands are designated as strand i, where i represents strand number; note that strands 9 and 10 belong to the back sheet. C and N represent the C and N termini of the protein, respectively.

(B) Proposed topological determinants in CRABP1. Shown in red backbone are the helix-turn-helix motif and turns III and IV—local sequence elements in CRABP1 found to adopt native-like structures in peptide models (Rotondi and Gierasch, 2003b; Sukumar and Gierasch, 1997). The major hydrophobic core is shown in orange spheres, and the minor core is shown in green spheres. C and N represent the C and N termini of the protein, respectively.

native structure as isolated fragments led to the proposal that these local sequences adopted structure early in folding, perhaps even in the unfolded ensemble. These features, in fact, have been predicted to act as folding-initiating nuclei in iLBPs (Nikiforovich and Frieden, 2002).

Two alternative mechanisms for the formation of native topology during iLBP folding can thus be envisioned: Either, as previously reported for the immunoglobulin fold (Cota et al., 2001; Fowler and Clarke, 2001; Hamill et al., 2000), residues participating in the conserved and highly networked major hydrophobic core act as a folding nucleus and specify the sheet topology, or, alternatively, local structural features that interact across the β-barrel closure region form early and restrict conformational space in a manner that is uniquely advantageous to the β-clam fold. Either way, the absence of iLBP-related misfolding diseases implies that their folding, whether driven by consolidation of the hydrophobic core or interaction of locally encoded structural features, is robust enough to bypass aggregation.

To distinguish between these two possibilities and uncover features that provide aggregation protection, we introduced single-residue substitutions at 33 sites in the CRABP1 sequence with extensive coverage of structural elements including the minor hydrophobic core near the barrel closure region. Observing the impact of these mutations on CRABP1 stability and unfolding kinetics revealed that its rate-determining transition state (TS) is highly polarized, with barrel closure interactions forming before the TS and interactions in the major hydrophobic core developing only after the TS. It is striking that barrel closure provides partner interactions for edge strand 10 and, in so doing, protects regions of high predicted aggregation propensity, including strands 3, 4, 9, and 10, which we have found to form the core of isolated CRABP1 aggregates. We speculate that this folding mechanism may be common among other iLBP family members and that it exemplifies an intrinsic aggregation-protection strategy.

RESULTS AND DISCUSSION

β-Barrel Closure Occurs before the Rate-Determining Transition State of CRABP1, but Packing of the Major Hydrophobic Core Occurs Posttransition State

To characterize the rate-determining TS of CRABP1, we used protein engineering methods (Fersht et al., 1992). Because CRABP1 folds via multiple states, we used the established strategy of comparing the effect of mutations on the unfolding barrier, $\Delta\Delta G_{‡-N}$, to their effect on native state stability, $\Delta\Delta G_{U-N}$ (assuming the unfolded state energy is negligibly perturbed), thus generating unfolding ϕ values (ϕ_U), to report on whether an interaction is present in the TS or only forms post-TS (Bueno et al., 2006; Connell et al., 2009; Huysmans et al., 2010). Since it is conventional to report ϕ values in the folding direction, we converted the ϕ_Us to ϕs ($\phi = 1 - \phi_U$). $\Delta\Delta G_{U-N}$ values for mutant forms of CRABP1 were extracted from two-state fits of equilibrium urea denaturation curves, and $\Delta\Delta G_{‡-N}$ values were obtained by linear extrapolation of plots of ln k_U versus [urea] to 0 M urea. Representative examples of equilibrium and kinetic data are shown in Figures 2A and 2B. The stabilities of the CRABP1 variants varied from 5.2 ± 0.2 kcal/mol for the most destabilized variant, Y133S, to 9.8 ± 0.1 kcal/mol for the minimally perturbed V58A variant (compare to 10.0 ± 0.2 kcal/mol for wild-type [WT] CRABP1) (see Table 1). All substitutions were destabilizing except for G70A in turn III ($\Delta\Delta G_{U-N}$ = 0.7 ± 0.2 kcal/mol); this variant was included in the analysis due to consistent ϕ values of the neighboring residues. The increased stability of G70A may arise because the dihedral angles of the backbone at this position are compatible with Ala, and the entropy cost is lower for folding to native with an Ala residue than Gly.

We did not anticipate significant perturbations in the denatured state of CRABP1 as a result of the substitutions except when the original residue was a Gly. In these cases, restricted conformational flexibility due to the introduction of a bulkier side chain may affect protein's denatured state ensemble. In order to assess the effect of substitutions, the ^1H-^{15}N heteronuclear single quantum coherence (HSQC) spectra for the Gly-to-Ala variants (G68A, G70A, and G78A) were compared to that of the WT protein under denaturing conditions. Only a few peaks were affected by the substitutions, as shown for the G78A variant in Figure S1 (available online), arguing that the ensemble of structures sampled by the chain is very similar.

The ϕ values for most of the residues belonging to the major intersheet hydrophobic core and adjacent turn III were between 0 and 0.3, consistent with specific packing of this region only after the TS (Figures 2A and 2B; Table 1). In a previous amide hydrogen/deuterium (H/D) exchange study, we showed that interstrand H bonds formed only during the rate-determining folding step of CRABP1 (Clark et al., 1997). The current data make it clear that specific side chain packing in the major hydrophobic core is also a late event in CRABP1 folding.

In contrast, residues in the minor hydrophobic core and along the strand 1′–strand 10 connection displayed high ϕ values (Figures 2A and 2B; Table 1), suggesting that side-chain-specific

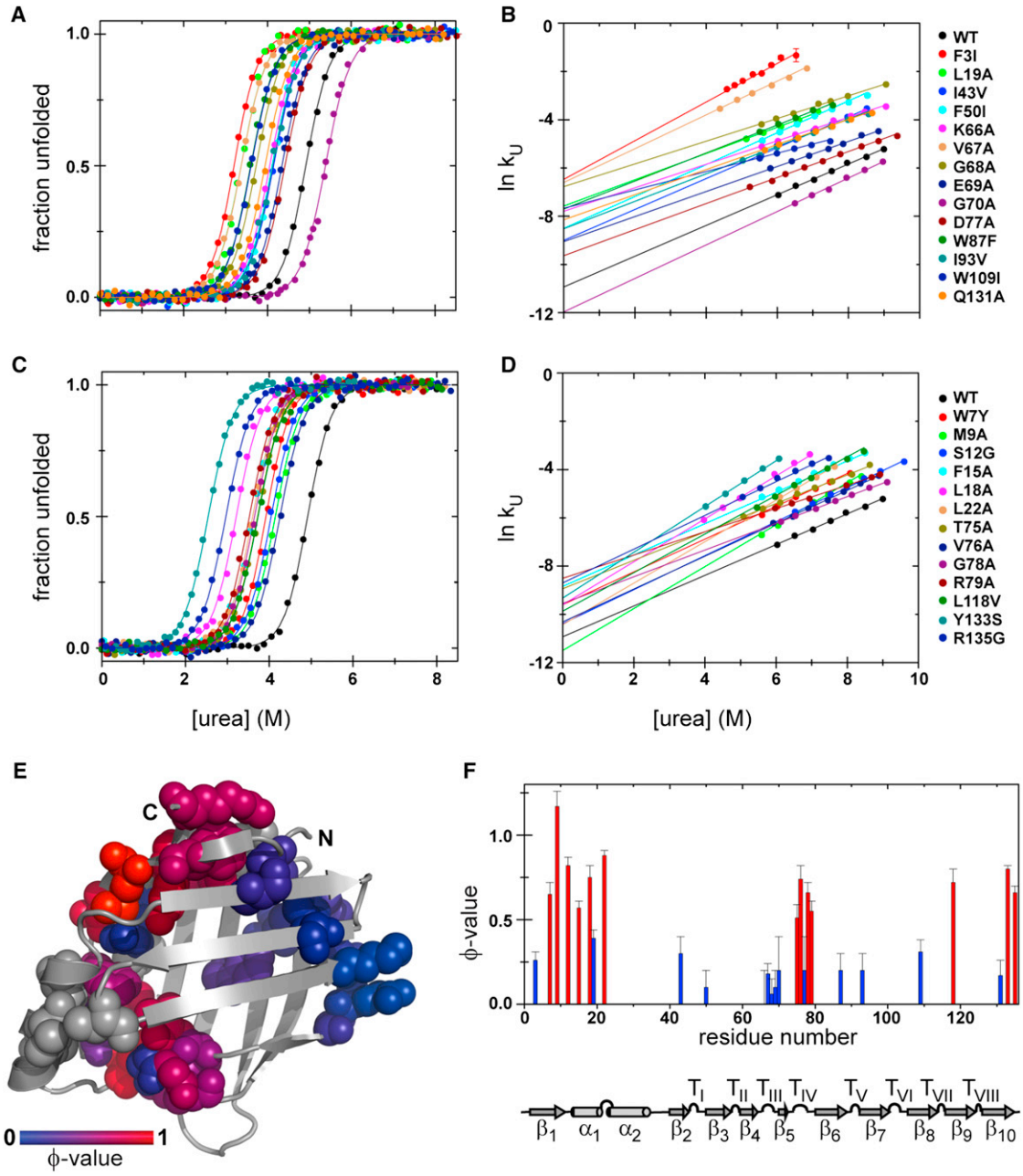

Figure 2. Equilibrium Stability, Unfolding Kinetics, and the Properties of the Rate-Determining TS of CRABP1

(A–D) In (A) and (C), representative curves are shown for urea-induced equilibrium denaturation for CRABP1 WT and its variants at 25°C. The solid lines are fits of the data to a two-state model using a fixed equilibrium m value of −2.0 kcal/mol × M. In (B) and (D), representative curves are shown for the dependence of the logarithm of the unfolding rates on urea for WT CRABP1 and its variants at 25°C; the solid lines on the ln k_U plots represent linear fits. (A) and (B) show the data for the residues that yield ϕ values < 0.5; (C) and (D) show the data for the residues with ϕ values > 0.5.

(E) The interactions present in the rate-determining TS mapped on to the native structure of CRABP1. Residues subjected to perturbation analysis are shown in spacefill and colored according to their ϕ values from blue (0) to red (1); those that caused no or minor perturbation are colored in gray.

(F) Sequence histogram with ϕ values for the CRABP1 variants under study. The secondary structural features of CRABP1 are shown below. Error bars represent SD from at least two independent measurements.

See also Figures S1 and S2.

interactions that are implicated in β-barrel closure form before the rate-determining TS. Residues showing high ϕ values include F15, L18, and L22 from helix I and V76, G78, and R79 from turn IV, the two local structural features previously demonstrated to form in peptide fragments (Rotondi and Gierasch, 2003b; Sukumar and Gierasch, 1997), as well as W7 from strand 1, M9 and S12 from strand 1′, L118 from strand 9, and Y133 and R135 from the C terminus of strand 10. The early involvement of W7

Table 1. Thermodynamic and Kinetic Parameters of CRABP1 Variants

Variant	Location[a]	$\Delta\Delta G_{U-N}$ (kcal/mol)[b]	ln k_U[c]	m_U (kcal/mol × M)[c]	$\Delta\Delta G_{‡-N}$ (kcal/mol)[d]	ϕ Value[e]
WT[f]		n/a	−11.0 ± 0.1	0.64 ± 0.02	n/a	n/a
F3I$^+$[g]	Strand 1	−3.6 ± 0.2	−6.5 ± 0.1	0.80 ± 0.03	−2.6 ± 0.1	0.26 ± 0.05
W7Y	Strand 1	−2.2 ± 0.2	−9.7 ± 0.2	0.68 ± 0.02	−0.8 ± 0.1	0.65 ± 0.07
M9A$^+$	Strand 1′	−1.8 ± 0.2	−11.5 ± 0.2	0.87 ± 0.02	0.3 ± 0.2	1.17 ± 0.09
S12G$^+$	Strand 1′	−1.8 ± 0.2	−10.38 ± 0.08	0.706 ± 0.008	−0.34 ± 0.08	0.82 ± 0.05
F15A	Helix I	−2.9 ± 0.2	−8.5 ± 0.2	0.65 ± 0.01	−1.25 ± 0.08	0.57 ± 0.04
L18A$^+$	Helix I	−3.5 ± 0.2	−9.5 ± 0.4	0.88 ± 0.05	−0.9 ± 0.2	0.75 ± 0.07
L19A$^-$[h]	Helix I	−3.3 ± 0.2	−7.58 ± 0.07	0.547 ± 0.005	−2.00 ± 0.09	0.39 ± 0.05
L22A$^+$	Helix I	−2.7 ± 0.2	−10.43 ± 0.02	0.859 ± 0.002	−0.31 ± 0.08	0.88 ± 0.03
R29A$^-$	Helix II	−0.8 ± 0.3	−9.17 ± 0.04	0.458 ± 0.003	−1.06 ± 0.08	n/a
V31A	Helix II	−0.6 ± 0.3	−10.6 ± 0.2	0.587 ± 0.008	−0.2 ± 0.1	n/a
V33A$^-$	Helix II	−0.4 ± 0.3	−9.9 ± 0.5	0.54 ± 0.05	−0.6 ± 0.3	n/a
I43V	Strand 2	−1.6 ± 0.2	−8.9 ± 0.1	0.61 ± 0.05	−1.2 ± 0.1	0.3 ± 0.1
F50I	Strand 3	−1.7 ± 0.2	−8.51 ± 0.04	0.66 ± 0.01	−1.45 ± 0.08	0.1 ± 0.1
V58A	Turn II	−0.2 ± 0.2	−10.9 ± 0.2	0.642 ± 0.002	0.0 ± 0.1	n/a
K66A$^-$	Turn III	−1.8 ± 0.3	−7.85 ± 0.07	0.48 ± 0.01	−1.84 ± 0.09	0.0 ± 0.2
V67A	Turn III	−3.3 ± 0.2	−6.5 ± 0.2	0.70 ± 0.03	−2.7 ± 0.1	0.18 ± 0.06
G68A$^-$	Turn III	−2.6 ± 0.2	−6.9 ± 0.2	0.50 ± 0.03	−2.4 ± 0.2	0.06 ± 0.09
E69A$^-$	Turn III	−1.3 ± 0.2	−9.02 ± 0.07	0.51 ± 0.01	−1.15 ± 0.09	0.1 ± 0.1
G70A	Turn III	0.7 ± 0.2	−11.9 ± 0.1	0.69 ± 0.01	0.6 ± 0.1	0.2 ± 0.2
T75A$^-$	Turn IV	−2.4 ± 0.3	−9.0 ± 0.1	0.54 ± 0.05	−1.2 ± 0.1	0.51 ± 0.08
V76A	Turn IV	−1.5 ± 0.3	−10.3 ± 0.1	0.675 ± 0.007	−0.4 ± 0.1	0.74 ± 0.08
D77A$^-$	Turn IV	−1.0 ± 0.3	−9.56 ± 0.07	0.529 ± 0.001	−0.83 ± 0.09	0.2 ± 0.2
G78A$^-$	Turn IV	−2.5 ± 0.4	−9.53 ± 0.04	0.547 ± 0.009	−0.85 ± 0.08	0.66 ± 0.06
R79A$^-$	Turn IV	−3.1 ± 0.3	−8.6 ± 0.2	0.46 ± 0.02	−1.4 ± 0.1	0.55 ± 0.06
W87F	Strand 6	−2.6 ± 0.4	−7.7 ± 0.1	0.59 ± 0.02	−1.9 ± 0.1	0.2 ± 0.1
I93V$^-$	Strand 7	−1.8 ± 0.2	−8.49 ± 0.05	0.567 ± 0.004	−1.46 ± 0.08	0.2 ± 0.1
W109I$^-$	Strand 8	−2.9 ± 0.2	−7.6 ± 0.1	0.37 ± 0.02	−2.0 ± 0.1	0.31 ± 0.07
L113V	Strand 8	−0.8 ± 0.2	−10.48 ± 0.02	0.66 ± 0.01	−0.28 ± 0.08	n/a
L118V$^+$	Strand 9	−2.3 ± 0.4	−9.9 ± 0.2	0.79 ± 0.03	−0.7 ± 0.2	0.72 ± 0.08
C129A$^-$	Strand 10	−0.6 ± 0.2	−9.8 ± 0.1	0.55 ± 0.01	−0.68 ± 0.09	n/a
Q131A$^-$	Strand 10	−2.0 ± 0.2	−8.15 ± 0.02	0.516 ± 0.006	−1.66 ± 0.08	0.17 ± 0.09
Y133S$^+$	Strand 10	−4.8 ± 0.2	−9.32 ± 0.02	0.950 ± 0.001	−0.97 ± 0.08	0.80 ± 0.02
R135G$^+$	Strand 10	−3.9 ± 0.3	−8.7 ± 0.1	0.71 ± 0.01	−1.34 ± 0.09	0.66 ± 0.04

The average values from at least two independent measurements and the SD are reported; all values reported are extrapolated to 0 M denaturant. n/a, not applicable.

[a]Secondary structural location of the selected variant residue.

[b]The energetic effect of mutations on the free energy of the unfolded state (U) with respect to the native state (N) calculated as $\Delta\Delta G_{U-N} = \Delta G_{U-N}^{WT} - \Delta G_{U-N}^{mut}$, where ΔG_{U-N}^{WT} was determined to be 10.0 ± 0.2 kcal/mol, and ΔG_{U-N} values were obtained by fitting equilibrium urea unfolding curves using $m_{eq} = -2.0$ kcal/mol × M (Figures 2A and 2C).

[c]Obtained from linear extrapolation of ln k_U versus [urea] (Figures 2B and 2D).

[d]The energetic effect of mutation on the free energy of the transition state (‡) with respect to the native state (N), calculated as $\Delta\Delta G_{‡-N} = -RT \ln (k_U^{WT}/k_U^{mut})$.

[e]Calculated as $\phi_F = 1 - \phi_U$.

[f]His-tagged CRABP1 with a stabilizing R131Q mutation was used as WT.

[g]Designates variants with $m_U > 0.7$ kcal/mol × M (m_U^+).

[h]Designates variants with $m_U < 0.57$ kcal/mol × M (m_U^-).

was suggested by our previous observation that its fluorescence changed in the dead time in stopped-flow mix refolding experiments (~3 ms; Clark et al., 1998). Altogether, residues with high ϕ values, indicating formation of their interactions before the rate-determining TS, comprise a structurally contiguous group in the native protein and participate in some of the longest range interactions in the native structure of CRABP1 (Figure 2A). Finally, mutations in helix II (R29A, V31A, V33A) and turn II (V58A)

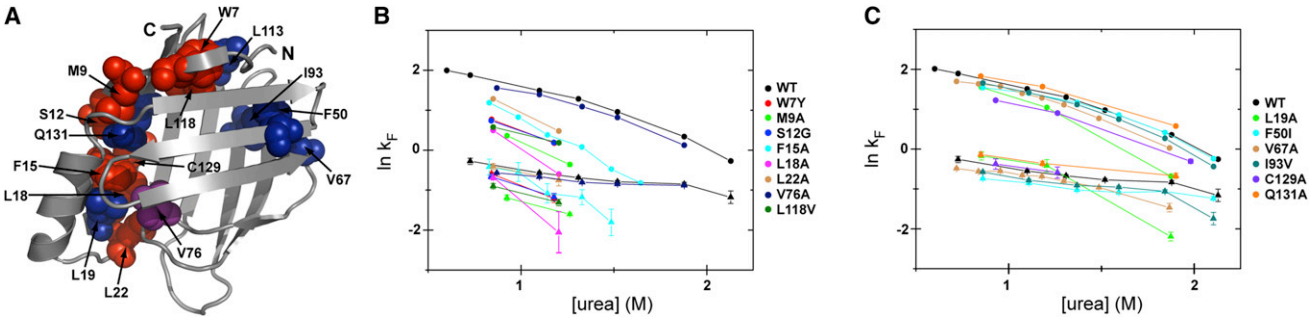

Figure 3. Refolding Kinetics for Selected CRABP1 Variants

(A) CRABP1 variants used in refolding kinetics experiments are shown on the CRABP1 structure: red spheres represent variants with ϕ values > 0.5 and significant retardation of refolding kinetics; blue spheres represent variants with ϕ values < 0.5 and WT-like refolding behavior. V76A, which has a ϕ value of 0.74 but WT-like stopped-flow refolding, is shown in violet; C129A, which perturbs the fast phase, is shown in red.

(B) Dependence of the logarithm of the rates of I_2 and N formation (fast and medium kinetic phases, respectively) on [urea] at 25°C for WT CRABP1 and selected variants with ϕ values > 0.5. Shown are plots of ln k_F versus [urea] for the fast (circles) and medium (triangles) phases. Error bars denote SEM.

(C) Same as in (B) for the CRABP1 variants with ϕ values < 0.5. Error bars denote SEM.

Colors correspond to those in Figures 1A–1D except for the C129A variant added in this figure.

See also Table S1.

minimally perturbed the native state stability and did not change CRABP1 unfolding kinetics (Table 1), despite formally contributing to the smaller hydrophobic core. This is consistent with the dynamic nature of helix II (Krishnan et al., 2000) and lack of native-like structural propensity in isolated turn II (Rotondi and Gierasch, 2003b) in CRABP1.

Overall, the rate-determining TS of CRABP1 is highly polarized with a folding nucleus encompassing helix I, turn IV, and strands 1, 1′, and 10 (Figures 2E and 2F) that is formed in the TS. This nucleus resembles the "two-strand-helix" nucleation motif shared by many α/β proteins (Lindberg and Oliveberg, 2007). It is important to note that, for CRABP1 β-barrel, this nucleation mechanism implies early barrel closure.

How Early in Folding Does Barrel Closure Occur?

To probe the timing of barrel closure, we examined the refolding kinetics of selected variants with substitutions in the barrel-closing region. Refolding of WT CRABP1 is triphasic, with the fastest phase (<3 ms) reporting on the formation of the hydrophobically collapsed I_1, followed by a fast phase (~200 ms) corresponding to formation of the topologically native but highly solvated intermediate I_2 and a 1 s phase corresponding to the formation of N (Clark et al., 1997, 1998). (Folding of a small population of CRABP1 molecules [about 20%] is limited by *cis-trans* isomerization of the L84-P85 bond, leading to a low-amplitude slow [15–20 s] phase [Eyles and Gierasch, 2000].) All CRABP1 variants with substitutions in the folding nucleus as determined by perturbation analysis (W7Y, M9A, S12G, F15A, L18A, L22A, and L118V; red spheres in Figure 3A) showed significantly retarded formation of both I_2 intermediate and N (Figure 3B; Table S1). The fact that they retained three refolding kinetic phases, like WT CRABP1, argues that they fold by the same mechanism, albeit with modest perturbation due to the mutations. The Y133S and R135G variants exhibited the slowest refolding kinetics (data not shown). Slower formation of I_2 implies structural and energetic perturbation of the TS barrier between I_1 and I_2. In an earlier study from our group, the fast phase in CRABP1 folding was attributed to docking of the N and C termini based on the fluorescence signature of W7 (Clark et al., 1998). The present data extend this model to implicate barrel closure, including docking of N and C termini, in the early stage of CRABP1 folding. It is noteworthy that the C129A substitution, which caused only minor destabilization of the native state ($\Delta\Delta G_{U-N}$ = −0.6 ± 0.2 kcal/mol) and did not formally "qualify" for ϕ value analysis, showed substantially retarded formation of I_2. Thus, C129 is most likely a peripheral part of the barrel-closing patch via its contacts with S12, F15, and L18. We speculate that early docking and formation of specific interactions between the terminal β strands may also help in simultaneous assembly of the front and the back sheets and thereby aid in rapid folding of the barrel.

Overall, analysis of the folding kinetics for residues with high ϕ values paints a consistent picture of barrel closure occurring between the I_1 and I_2 intermediates, well before the rate-determining TS. It is intriguing that one class of residue that presents a high ϕ value deviates from the pattern described earlier: The folding phases that are detectable by stopped-flow mixing kinetics experiments were essentially unperturbed from WT for turn IV variants such as V76A (Figure 3B; Table S1; also T75A, G78A, and R79A at 0.73 M urea). We conclude that turn IV formation occurs very early. In fact, the corresponding backbone region may sample turn conformations in the unfolded state as suggested by previous peptide studies (Rotondi and Gierasch, 2003b), while specific interactions between side chains may occur in the I_1. Indeed, another study on the turn IV peptide model has shown that V76A, D77A, and R79A substitutions did not fully reduce turn propensity (Rotondi and Gierasch, 2003a). Experiments are underway to test this possibility in the context of the full-length CRABP1. Note that turn IV has two key roles in folding: It joins residues from helix II and strands 1 and 1′ to form the β-barrel closure cluster, and it is in a topologically strategic position splitting front and back sheets via a β bulge.

The primary perturbation associated with residues that display low ϕ values is expected to be destabilization of the native state, with little change in folding kinetics. Consistent with their ϕ values, CRABP1 variants F50I, V67A, I93V, L113V, and Q131A

(blue spheres in Figure 3A) showed essentially unperturbed WT-like refolding behavior (Figure 3C; Table S1). Thus, our data strongly argue for the formation of specific side chain interactions in the barrel-closing region as early as in I_1 intermediate, while the rest of the protein including helix II and most of the major conserved hydrophobic core remains plastic until after the rate-determining TS.

The Rate-Determining Transition State of CRABP1 Is Malleable

Several CRABP1 variants exhibited unfolding m values (m_U values) more than 10% different from that of the WT protein (0.64 kcal/mol × M). These variants were classified as m_U^+ ($m_U > 0.7$ kcal/mol × M) or m_U^- ($m_U < 0.57$ kcal/mol × M). The dependence of the difference between free energies of the native and the denatured states on [denaturant] is proportional to the accessible surface area (ASA) change between these states (Myers et al., 1995); the dependence of the logarithm of the unfolding rate constant ($\ln k_U$) on [denaturant] is proportional to the ASA change between the native state and the rate-determining TS. Generally, the change in m_U may be explained in terms of movement of the TS, perturbation of the native state, or change of the folding pathway (Fowler and Clarke, 2001; Matouschek and Fersht, 1993; Otzen and Oliveberg, 2002). Since refolding behavior of all CRABP1 variants tested was triphasic similar to the WT, we favor the interpretation that it is the movement of the rate-determining TS that accounts for the observed m_U changes. The observation of unperturbed equilibrium m values strongly supports this conclusion, suggesting that, for all variants used in this study, both the denatured and native states resemble WT CRABP1 in terms of ASA, as simultaneous and equal perturbation of ASA of both states is unlikely. Therefore, the altered m_U values likely report on the changes in the ASA of the corresponding TS, rather than on the native state. Indeed, for both m_U^+ and m_U^- variants, we observed curvature in plots of the log of the unfolding rate constant as a function of [urea] at high denaturant (Figure S2), thus indicating that the rate-determining TS of CRABP1 is also malleable with respect to the denaturant. We were unable to discern any curvature in the log unfolding rate plot of WT protein over the experimentally accessible urea range. The stronger denaturant dependence of the unfolding rate for the m_U^+ variants argues that the structure of their TS is expanded relative to that of the WT CRABP1. Conversely, the m_U^- variants have relatively more compact TS. We cannot, however, rule out selective destabilization of the native state by some mutations, which would lead to the accumulation of an expanded native-like intermediate in the burst phase of the unfolding reaction (Otzen and Oliveberg, 2002; Seeliger et al., 2003).

It is interesting that the distribution of ϕ values qualitatively correlates with that of m_U values. Seven out of eight m_U^+ variants (M9A, S12G, L18A, L22A, L118V, Y133S, and R135G) are in the early folding core. The expanded TS of these variants can be explained by looser packing due to substitution of a residue with a bulky side chain by one with a small side chain at critical positions where specific side chain packing in the TS occurs. Most of the m_U^- variants are polar-to-nonpolar substitutions (R29A, K66A, E69A, T75A, D77A, R79A, W109I, C129A, and Q131A). Here, removal of electrostatic interactions and introduction of a hydrophobic side chain may lead to tighter, but non-specific packing, which results in compaction of the TS. The same argument holds for G68A and G78A substitutions, which increased side chain hydrophobicity. The L19A, V33A, and I93V variants are in regions that lack specific side chain packing; these m_U^- substitutions mirror the effects of similar m_U^+ ones in the early folding core.

This deeper analysis of the effects of substitutions on denaturant dependence of unfolding kinetics reveals intriguing aspects of the TS ensemble and its packing properties. The rate-determining TS is at once malleable, as substitutions alter its compaction/expansion, and robust, in that all of our observations can be interpreted in terms of a single self-consistent TS ensemble.

β-Barrel Closure Protects Aggregation-Prone Regions of CRABP1

While no iLBP has been implicated in a misfolding disease, CRABP1 and variants aggregate to varying extents in vitro or in *E. coli* expression systems, indicating that aggregation of this protein indeed competes with its folding. Multiple sequence-based aggregation prediction algorithms, including Zyggregator (Tartaglia and Vendruscolo, 2008) and PASTA (Trovato et al., 2007), consistently identify several sequence stretches in the β strands of CRABP1 as aggregation prone (Figure 4A). To test these predictions experimentally, we identified sequences that form the core of CRABP1 aggregates using nuclear magnetic resonance (NMR) spectroscopy (Hoshino et al., 2002). Specifically, we determined the amide H/D exchange profile of in vitro-prepared aggregates of a highly aggregation-prone CRABP1 variant (F71A). The overlay of ^1H-^{15}N HSQC spectra of CRABP1 F71A aggregates dissolved in DMSO before and after four weeks of incubation in D_2O is shown in Figure 4B. Consecutive stretches including residues 51-65, which corresponds to strand 3-turn II-strand 4 with the exception of S55, residues 119–123 (strand 9), and 127–134 (strand 10) are highly protected even after 4 weeks of exchange (magenta bars in Figure 4A and regions on structure in Figure 4C). We conclude that these 27 residues constitute the core of CRABP1 aggregates. These experimentally determined core sequences agree well with the Zyggregator (Tartaglia and Vendruscolo, 2008) and PASTA (Trovato et al., 2007) predictions. It is noteworthy that strand 10 belongs to the aggregation core. This strand can be categorized as an "edge strand" in terms of the architecture of CRABP1. Early barrel closure in the folding mechanism of CRABP1 would protect strand 10 by providing it a set of partner interactions, thus mitigating its vulnerability as an unpartnered "edge strand" (Richardson and Richardson, 2002). Additionally, inspection of the location of the high ϕ residues in the CRABP1 folding mechanism and the aggregation-prone regions suggests a more general protection of these regions by the early closure of the barrel. For example, intrinsically labile strand 9 (Krishnan et al., 2000; Xiao and Kaltashov, 2005) also benefits from protection, which, in this case, may be provided by strand 10 once barrel closure occurs. Thus, we hypothesize that protection of aggregation-prone regions in CRABP1 by structural features and barrel closure occurs early in folding and significantly reduces its risk of aggregation. In other words, the folding pathway is driven by topology so as to avoid misfolding, rather

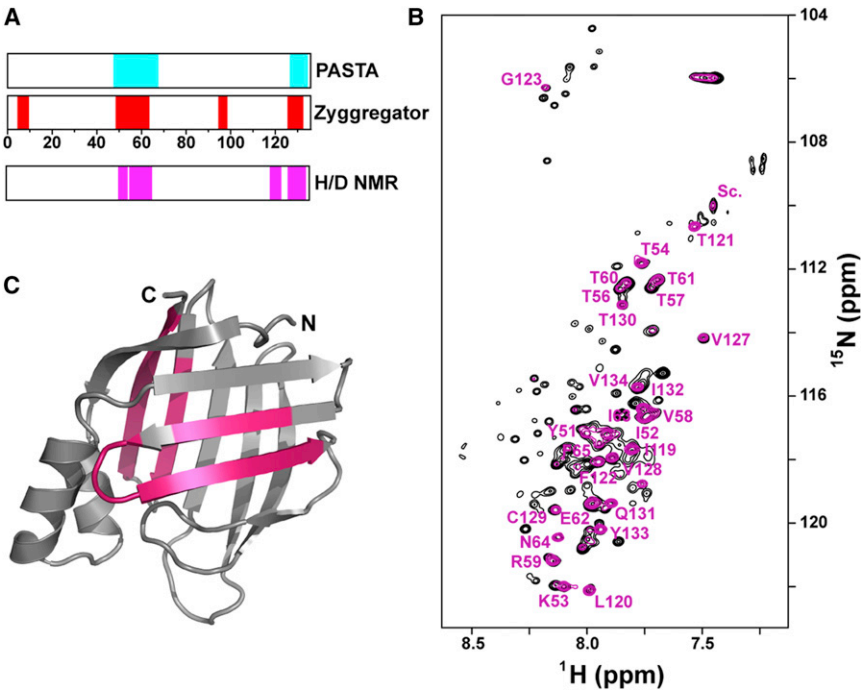

Figure 4. Predicted and Experimentally Identified Aggregation-Prone Regions in CRABP1

(A) Regions of CRABP1 predicted to be aggregation prone by the Zyggregator (Tartaglia and Vendruscolo, 2008) and PASTA (Trovato et al., 2007) algorithms shown in red and cyan bars, respectively. Magenta bars indicate the actual aggregation cores determined by NMR.

(B) ^{1}H-^{15}N HSQC spectra of CRABP1 F71A aggregates dissolved in 95% d_6-DMSO/5% D_2O. Resonances in black correspond to unexchanged aggregates, while resonances in magenta correspond to aggregates that have been H/D-exchanged for 4 weeks. Peaks that retained significant intensity in the exchanged sample are identified.

(C) Regions of CRABP1 F71A found to be protected in the H/D exchange experiment (less than 50% drop in concentration-normalized peak intensity) are illustrated on the structure of CRABP1.

See also Figure S3.

than by a specific hydrophobic collapse, which was the previously accepted folding mechanism in this structural family. Furthermore, these results provide direct experimental evidence that connects the folding and aggregation landscapes of a protein.

Is Early Barrel Closure a General Feature of iLBP Folding?

Can our findings on CRABP1 be extrapolated to other iLBP family members or other protein structural classes? While data for other iLBPs are not as extensive as those available for CRABP1, there is some evidence supporting the generality of early barrel closure as a feature on iLBP folding landscapes. First, most of the residues identified in this study as parts of the folding nucleus of CRABP1 are highly conserved among iLBPs (Figure 5A). Moreover, the majority of the highly conserved residues in iLBPs are localized to two spatially contiguous structural regions that define a β-barrel topology: (1) the turn III-strand five-turn IV element that connects the two sheets and (2) barrel-closing strands 1 and 10. In addition, the very C terminus has been shown to be a critical topological determinant. Deletion of three C-terminal residues in rat intestinal fatty acid-binding protein (IFABP) produced a swollen molten globule with no specific aromatic tertiary interactions and little residual secondary structure (Clérico et al., 2000). Analogous deletion in A CRABP1 background also yields a molten globule-like nonnative state (B.K. and E.M. Clérico, unpublished data). Substitution of the helix-turn-helix motif with a short linker turns causes complete loss of native structure in the apo-form of rabbit ileal lipid binding protein (ILBP) (Kouvatsos et al., 2007). It is somewhat surprising that removal of the helical domain from rat IFABP is tolerated but significantly retards refolding (Kim et al., 1996). More recent folding studies on rat IFABP suggest early formation of native-like structure around F2 and F17 (corresponding to F3 and L18 in CRABP1) (Li and Frieden, 2007), consistent with our findings on CRABP1. The folding pathway of rat IFABP deduced from equilibrium denaturation NMR experiments (Hodsdon and Frieden, 2001; Ropson and Frieden, 1992) provides additional parallels. Proposed initial folding steps for IFABP involve structure consolidation around the turn between two helices, turn III and the turn between strands 9 and 10. Similarly, a number of the main hydrophobic core residues in the back sheet were found to participate in the IFABP equilibrium intermediate observed between 5.0 and 6.0 M urea (Hodsdon and Frieden, 2001). This intermediate is likely to represent a submillisecond burst phase kinetic intermediate observed for CRABP1 and involves nonspecific collapse of the main hydrophobic core. It is important to note here that perturbation analysis data report on specific side chain interactions, while equilibrium denaturation NMR experiments report on the environment of amide backbones; the comparison between them may not always be straightforward (Zarrine-Afsar et al., 2012). This may explain why the authors have not observed early formation of turn IV but did observe that of turn III in IFABP. Taken together, our results and evidence from previous studies on iLBPs strongly argue that early barrel closure is a mechanistic step likely to be shared by several iLBP family members. Prediction of aggregation-prone regions of other members of iLBP family (Figure 5B) identifies similar regions to those predicted for CRABP1 and suggests that the early closure of the barrel could indeed be a general mechanism to increase the probability of productive folding.

Folding and aggregation landscapes are inevitably overlapping (Clark, 2004; Hartl et al., 2011; Jahn and Radford, 2008; Vendruscolo, 2012). Proteins have evolved to avoid aggregation through the stabilization of the native state and/or destabilization

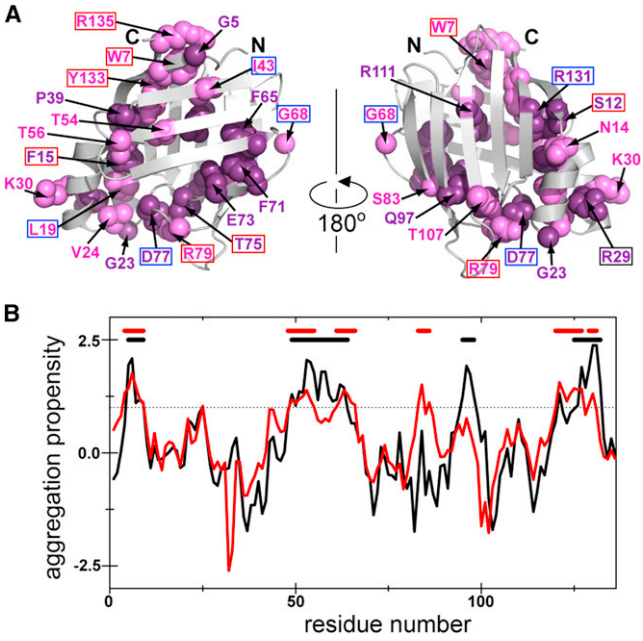

Figure 5. Residues that Form Contacts Early in the Folding Pre-TS and Constitute Aggregation-Prone Regions in CRABP1 Are Evolutionarily Conserved in the iLBP Family

(A) Sequence conservation residues in the iLBP family. Multiple sequence alignment of iLBPs to CRABP1 was carried out using the ConSurf server (Ashkenazy et al., 2010). Residues shown in dark purple spheres on the CRABP1 structure have the highest ConSurf score (9), and those shown in light purple have the next highest (8). Residue numbering corresponds to that in the CRABP1 sequence. Boxed residues were subjected to perturbation analysis in this study, and the box color corresponds to ϕ values (red for $\phi > 0.5$ and blue for $\phi < 0.5$).

(B) Aggregation propensity profiles for CRABP1 and iLBP family average predicted by Zyggregator (Tartaglia and Vendruscolo, 2008). The aggregation propensity profile for WT CRABP1 is shown in black, and the average iLBP aggregation profile in red. The horizontal dotted line represents the 1.0 threshold for significant aggregation propensity; continuous stretches of residues with aggregation scores above the threshold are considered aggregation prone and marked as black and red horizontal bars, respectively.

of risky intermediates (Chiti and Dobson, 2009), presence of residual structure in the aggregation-prone segments in the unfolded state ensemble (Hamada et al., 2009; Routledge et al., 2009), or even functional protein-protein interactions (Masino et al., 2011). Another strategy is implied by our studies of CRABP1, where it appears that folding steps may protect aggregation-prone regions. This mechanism is likely to be general, even beyond iLBPs and other β-barrel proteins. We note a couple of examples other than the iLBPs. Cold shock proteins (Csps) adopt the five-stranded β-barrel oligonucleotide/oligosaccharide binding fold (OB-fold). While strands 1 and 3 of CspA are implicated in its aggregation (Alexandrescu and Rathgeb-Szabo, 1999), the TS of the related CspB has been mapped and reveals that strands 1 and 4 are involved in pre-TS nonlocal specific interactions (Garcia-Mira et al., 2004). Linking these two studies using the assumption that the folding mechanism of CspA resembles that of CspB suggests that early folding events may act to diminish risk of aggregation in this family as well. Additionally, a surface loop in interleukin 1β has been proposed to form a hydrophobically structured microdomain (Chrunyk and Wetzel, 1993) and speculated to pack against strands 6–10 to prevent self-association (Finke et al., 2000), while terminal strands remain exposed until barrel closure late in folding (Chavez et al., 2006). There are counter examples: For example, a recent study of the FynSH3 domain elegantly described an aggregation precursor in which the carboxyl terminal β strand 5 is unstructured (Neudecker et al., 2012). This disrupts its interactions with β strand 1, which is an aggregation-prone region. However, based on earlier studies (Northey et al., 2002), neither of these terminal strands exhibits specific interactions in the rate-determining TS. Thus, the folding mechanism in this case does not appear to protect against aggregation via strand 1. In addition, the aggregation-prone intermediate is only sparsely populated, suggesting that there are other strategies in play to minimize aggregation. Thus, we suspect that in-depth studies of additional proteins will reveal that there are indeed multiple strategies that arise from the evolutionary pressures on folding mechanisms to suppress the likelihood of aggregation.

EXPERIMENTAL PROCEDURES

Protein Design, Expression, and Purification

A variant of murine CRABP1 with an N-terminal $(His)_{10}$-tag and a stabilizing R131Q mutation (Clark et al., 1998), referred to as WT (CRABP1 WT) here, was used as a template for mutagenesis. The single-site variants were generated by site-directed mutagenesis using a QuikChange protocol (Stratagene) and confirmed by DNA sequencing. The WT and mutant proteins were expressed in *E. coli* strain BL21(DE3) (Novagen). In the case of variants, 30 min prior to induction, L-Pro and NaCl were added to the growing culture to a final concentration of 20 mM and 0.3 M, respectively, to improve protein solubility (Ignatova and Gierasch, 2006). Protein expression was induced with 0.4 mM IPTG, and the cells were allowed to grow for 4 hr at 30°C. All proteins were purified upon cell lysis from the soluble fraction of the cell extract by Ni-NTA (QIAGEN) affinity chromatography. Protein concentration was determined using a molar extinction coefficient of ε_{280} = 20,970 $M^{-1}cm^{-1}$. The mutations were also validated by electrospray ionization mass spectrometry of pure proteins.

Equilibrium Denaturation

Protein samples (about 5 μM protein concentration) in 10 mM Tris-HCl, pH 8.0, containing 1 mM dithiothreitol (DTT) and varying urea concentrations were equilibrated overnight (16–18 hr) at 25°C. Unfolding transitions were monitored by Trp fluorescence (excitation at 280 nm, emission at 350 nm). The data were analyzed by a two-state model using the linear extrapolation method (Pace, 1986). The equilibrium m values (m_{eq}) for all variants were within ±10% of that of WT CRABP1. Therefore, the m_{eq} was fixed to −2.0 kcal/mol × M to reduce errors in the determination of ΔG_{U-N}.

Unfolding Kinetics

Protein samples in 10 mM Tris-HCl, pH 8.0, 1 mM DTT were manually added to urea solutions in the same buffer to a final concentration of about 5 μM. Unfolding kinetics was monitored by following Trp fluorescence (excitation at 280 nm, emission at 350 nm) at 25°C. The rate constant of unfolding (k_U) was determined from data fitted to a single exponential equation with a linear component introduced to account for photobleaching.

Refolding Kinetics

Refolding kinetics was followed by Trp fluorescence (excitation at 280 nm, emission at 350 nm) using an SFM-400 stopped-flow device (BioLogic) at 25°C. Urea-denatured protein (about 100 μM) in 20 mM Tris-HCl, pH 8.0, 1 mM DTT was refolded by dilution into refolding buffer (20 mM Tris-HCl, pH 8.0, 1 mM DTT containing varying urea concentrations). A cuvette with a pathlength of 0.8 mm was used. The dead time of the instrument was

2.4 ms. Kinetics traces were averaged and fit to a multiexponential equation using Origin (OriginLab) or SigmaPlot (Systat Software).

Perturbation Analysis

Perturbation analysis, also known as ϕ value analysis (Fersht et al., 1992), was performed on 33 variants of CRABP1. Typically, ϕ values report on the onset of specific interactions with respect to the rate-determining TS. Unfolding ϕ values (ϕ_U values) were calculated as a ratio between the energetic perturbation of the TS ($\Delta\Delta G_{I-N}$) and the native state ($\Delta\Delta G_{U-N}$) upon mutation. Only those mutations that caused significant perturbation ($\Delta\Delta G_{U-N} > 0.8$ kcal/mol) were considered for analyses. ϕ_U values were converted to the folding ϕ_F values (ϕ) as $\phi = 1 - \phi_U$. Accordingly, $\phi = 0$ was interpreted as an indication of formation of residue-specific native-like interactions after the rate-determining TS, and $\phi = 1$ was interpreted as an indication of formation of those before the rate-determining TS. Note that all ϕ values reported in the text are folding ϕ_F values and referred to as ϕ values for the sake of simplicity.

Preparation of CRABP1 Aggregates

The $^{15}N/^{13}C$-labeled F71A variant of WT CRABP1 was overexpressed in BL21(DE3) *E. coli* cells, and the protein was purified from the inclusion bodies as reported previously (Clark et al., 1998) with minor modification. Cells were lysed, and the pellet fraction was dissolved in 300 mM NaCl, 50 mM NaP_i, pH 8.0, containing 8 M urea. The protein was applied to a Ni-NTA agarose column in 8 M urea in 300 mM NaCl, 20 mM Tris-HCl, pH 8.0, and eluted using an imidazole gradient (from 50 mM to 300 mM). Aggregates were prepared by overnight dialysis of the urea-denatured protein against 150 mM NaCl, 10 mM NaP_i, pH 7.0, 5 mM DTT at 37°C. Aggregates formed upon dialysis were collected by centrifugation and resuspended in D_2O containing 0.025% (w/v) NaN_3.

H/D Exchange Measurements

The regions of CRABP1 aggregates that constitute its aggregation core were identified by H/D exchange monitored by solution NMR spectroscopy as described previously (Hoshino et al., 2002) with minor modifications. Briefly, in vitro aggregates of $^{15}N/^{13}C$-labeled CRABP1 F71A were resuspended in D_2O. After 4 weeks of exchange at 4°C, aggregates were collected, lyophilized, and then resuspended in d_6-DMSO containing 0.1% (v/v) trifluoroacetic acid, 50 mM DTT, and 5% (v/v) D_2O (pD 3.0) to a final protein concentration of 200 μM. The solution was immediately transferred to an NMR tube, and the HSQC spectrum was recorded at 26°C on a 600 MHz Bruker Avance spectrometer using a TXI cryoprobe. For unexchanged samples, aggregates were resuspended in water containing 0.025% (w/v) NaN_3 and incubated at 4°C for the same amount of time. Data were processed using NMRPipe (Delaglio et al., 1995) and CARA (Keller, 2004). Backbone assignments of CRABP1 aggregates dissolved in DMSO were obtained using a standard set of triple resonance experiments, including HNCACB, CBCA(CO)NH, HBHANH, HNCO, and HNCACO.

Sequence Analysis

Conservation analysis on the iLBP family was performed using the ConSurf server (Ashkenazy et al., 2010) on 263 unique sequences out of 469 PSI-BLAST hits with murine CRABP1 as a query. For aggregation propensity analysis in the iLBP family, the sequences were retrieved by ExPasy BLAST (Artimo et al., 2012) again using murine CRABP1 as a query. Data set redundancy was decreased by filtering for sequences with 30%–90% identity and with no significant gaps and/or insertions. The sequences were submitted to the Zyggregator server to predict aggregation propensities for each iLBP. The aggregation propensity score was calculated for each position corresponding to the WT CRABP1 sequence and normalized by the total number of residues occurring at corresponding position according to a MAFFT alignment (Katoh and Toh, 2008).

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.01.013.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (OD000945).

Received: October 1, 2012
Revised: December 19, 2012
Accepted: January 15, 2013
Published: February 28, 2013

REFERENCES

Alexandrescu, A.T., and Rathgeb-Szabo, K. (1999). An NMR investigation of solution aggregation reactions preceding the misassembly of acid-denatured cold shock protein A into fibrils. J. Mol. Biol. *291*, 1191–1206.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. *40*(Web Server issue), W597–W603.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*(Web Server issue), W529–W533.

Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S., and Jones, T.A. (1994). Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. Adv. Protein Chem. *45*, 89–151.

Bueno, M., Ayuso-Tejedor, S., and Sancho, J. (2006). Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. J. Mol. Biol. *359*, 813–824.

Chavez, L.L., Onuchic, J.N., and Clementi, C. (2004). Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. J. Am. Chem. Soc. *126*, 8426–8432.

Chavez, L.L., Gosavi, S., Jennings, P.A., and Onuchic, J.N. (2006). Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. Proc. Natl. Acad. Sci. USA *103*, 10254–10258.

Chiti, F., and Dobson, C.M. (2009). Amyloid formation by globular proteins under native conditions. Nat. Chem. Biol. *5*, 15–22.

Chrunyk, B.A., and Wetzel, R. (1993). Breakdown in the relationship between thermal and thermodynamic stability in an interleukin-1 beta point mutant modified in a surface loop. Protein Eng. *6*, 733–738.

Clark, P.L. (2004). Protein folding in the cell: reshaping the folding funnel. Trends Biochem. Sci. *29*, 527–534.

Clark, P.L., Liu, Z.P., Rizo, J., and Gierasch, L.M. (1997). Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. Nat. Struct. Biol. *4*, 883–886.

Clark, P.L., Weston, B.F., and Gierasch, L.M. (1998). Probing the folding pathway of a beta-clam protein with single-tryptophan constructs. Fold. Des. *3*, 401–412.

Clérico, E.M., Peisajovich, S.G., Ceolín, M., Ghiringhelli, P.D., and Ermácora, M.R. (2000). Engineering a compact non-native state of intestinal fatty acid-binding protein. Biochim. Biophys. Acta *1476*, 203–218.

Connell, K.B., Miller, E.J., and Marqusee, S. (2009). The folding trajectory of RNase H is dominated by its topology and not local stability: a protein engineering study of variants that fold via two-state and three-state mechanisms. J. Mol. Biol. *391*, 450–460.

Cota, E., Steward, A., Fowler, S.B., and Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. J. Mol. Biol. *305*, 1185–1194.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Eyles, S.J., and Gierasch, L.M. (2000). Multiple roles of prolyl residues in structure and folding. J. Mol. Biol. *301*, 737–747.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. *224*, 771–782.

Finke, J.M., Roy, M., Zimm, B.H., and Jennings, P.A. (2000). Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1beta. Biochemistry *39*, 575–583.

Fowler, S.B., and Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. Structure *9*, 355–366.

Garcia-Mira, M.M., Boehringer, D., and Schmid, F.X. (2004). The folding transition state of the cold shock protein is strongly polarized. J. Mol. Biol. *339*, 555–569.

Gunasekaran, K., Hagler, A.T., and Gierasch, L.M. (2004). Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. Proteins *54*, 179–194.

Hamada, D., Tanaka, T., Tartaglia, G.G., Pawar, A., Vendruscolo, M., Kawamura, M., Tamura, A., Tanaka, N., and Dobson, C.M. (2009). Competition between folding, native-state dimerisation and amyloid aggregation in beta-lactoglobulin. J. Mol. Biol. *386*, 878–890.

Hamill, S.J., Steward, A., and Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. J. Mol. Biol. *297*, 165–178.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324–332.

Hodsdon, M.E., and Frieden, C. (2001). Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. Biochemistry *40*, 732–742.

Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H., and Goto, Y. (2002). Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. Nat. Struct. Biol. *9*, 332–336.

Huysmans, G.H., Baldwin, S.A., Brockwell, D.J., and Radford, S.E. (2010). The transition state for folding of an outer membrane protein. Proc. Natl. Acad. Sci. USA *107*, 4099–4104.

Ignatova, Z., and Gierasch, L.M. (2006). Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. Proc. Natl. Acad. Sci. USA *103*, 13357–13361.

Jahn, T.R., and Radford, S.E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. Arch. Biochem. Biophys. *469*, 100–117.

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298.

Keller, R. (2004). Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment (Zurich, Switzerland: Swiss Federal Institute of Technology).

Kim, K., Cistola, D.P., and Frieden, C. (1996). Intestinal fatty acid-binding protein: the structure and stability of a helix-less variant. Biochemistry *35*, 7553–7558.

Kleywegt, G.J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., and Jones, T.A. (1994). Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. Structure *2*, 1241–1258.

Kouvatsos, N., Thurston, V., Ball, K., Oldham, N.J., Thomas, N.R., and Searle, M.S. (2007). Bile acid interactions with rabbit ileal lipid binding protein and an engineered helixless variant reveal novel ligand binding properties of a versatile beta-clam shell protein scaffold. J. Mol. Biol. *371*, 1365–1377.

Krishnan, V.V., Sukumar, M., Gierasch, L.M., and Cosman, M. (2000). Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding. Biochemistry *39*, 9119–9129.

Li, H., and Frieden, C. (2007). Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. Proc. Natl. Acad. Sci. USA *104*, 11993–11998.

Lindberg, M.O., and Oliveberg, M. (2007). Malleability of protein folding pathways: a simple reason for complex behaviour. Curr. Opin. Struct. Biol. *17*, 21–29.

Marcelino, A.M., Smock, R.G., and Gierasch, L.M. (2006). Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. Proteins *63*, 373–384.

Masino, L., Nicastro, G., Calder, L., Vendruscolo, M., and Pastore, A. (2011). Functional interactions as a survival strategy against abnormal aggregation. FASEB J. *25*, 45–54.

Matouschek, A., and Fersht, A.R. (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. Proc. Natl. Acad. Sci. USA *90*, 7814–7818.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. *4*, 2138–2148.

Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundström, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., and Kay, L.E. (2012). Structure of an intermediate state in protein folding and aggregation. Science *336*, 362–366.

Nikiforovich, G.V., and Frieden, C. (2002). The search for local native-like nucleation centers in the unfolded state of beta -sheet proteins. Proc. Natl. Acad. Sci. USA *99*, 10388–10393.

Northey, J.G., Di Nardo, A.A., and Davidson, A.R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. Nat. Struct. Biol. *9*, 126–130.

Otzen, D.E., and Oliveberg, M. (2002). Conformational plasticity in folding of the split beta-alpha-beta protein S6: evidence for burst-phase disruption of the native state. J. Mol. Biol. *317*, 613–627.

Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. *131*, 266–280.

Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. J. Mol. Biol. *277*, 985–994.

Richardson, J.S., and Richardson, D.C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. USA *99*, 2754–2759.

Ropson, I.J., and Frieden, C. (1992). Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by 19F NMR. Proc. Natl. Acad. Sci. USA *89*, 7222–7226.

Rotondi, K.S., and Gierasch, L.M. (2003a). Local sequence information in cellular retinoic acid-binding protein I: specific residue roles in beta-turns. Biopolymers *71*, 638–651.

Rotondi, K.S., and Gierasch, L.M. (2003b). Role of local sequence in the folding of cellular retinoic acid binding protein I: structural propensities of reverse turns. Biochemistry *42*, 7976–7985.

Routledge, K.E., Tartaglia, G.G., Platt, G.W., Vendruscolo, M., and Radford, S.E. (2009). Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. J. Mol. Biol. *389*, 776–786.

Seeliger, M.A., Breward, S.E., and Itzhaki, L.S. (2003). Weak cooperativity in the core causes a switch in folding mechanism between two proteins of the cks family. J. Mol. Biol. *325*, 189–199.

Sukumar, M., and Gierasch, L.M. (1997). Local interactions in a Schellman motif dictate interhelical arrangement in a protein fragment. Fold. Des. *2*, 211–222.

Tartaglia, G.G., and Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. Chem. Soc. Rev. *37*, 1395–1401.

Trovato, A., Seno, F., and Tosatto, S.C. (2007). The PASTA server for protein aggregation prediction. Protein Eng. Des. Sel. *20*, 521–523.

Vendruscolo, M. (2012). Proteome folding and aggregation. Curr. Opin. Struct. Biol. *22*, 138–143.

Xiao, H., and Kaltashov, I.A. (2005). Transient structural disorder as a facilitator of protein-ligand binding: native H/D exchange-mass spectrometry study of cellular retinoic acid binding protein I. J. Am. Soc. Mass Spectrom. *16*, 869–879.

Zarrine-Afsar, A., Dahesh, S., and Davidson, A.R. (2012). A residue in helical conformation in the native state adopts a β-strand conformation in the folding transition state despite its high and canonical Φ-value. Proteins *80*, 1343–1349.